DOCUMENT RESUME

ED 428 237 CE 078 271

| | |
|---|---|
| TITLE | Performance Evaluation Process. |
| PUB DATE | 1998-03-00 |
| NOTE | 31p.; In: Academy of Human Resource Development Conference Proceedings (Oak Brook, Illinois, March 4-8, 1998); see CE 078 258. |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | EDRS Price MF01 Plus Postage. PC Not Available from EDRS. |
| DESCRIPTORS | Adult Education; Case Studies; Educational Assessment; *Evaluation Methods; Foreign Countries; Insurance Companies; *Job Performance; *Labor Force Development; *On the Job Training; *Personnel Evaluation; Program Effectiveness; Relationship; Teamwork; Technical Occupations; Total Quality Management |
| IDENTIFIERS | Netherlands |

ABSTRACT

This document contains four papers from a symposium on the performance evaluation process and human resource development (HRD). "Assessing the Effectiveness of OJT (On the Job Training): A Case Study Approach" (Julie Furst-Bowe, Debra Gates) is a case study of the effectiveness of OJT in one of a high-tech manufacturing company's product lines. Two studies of the nature and extent of insurance industry training of property/casualty and life/health agents are examined in "Linking HRD to Agents' Performance in the Insurance Industry" (Teresa M. Palmer, Kenneth J. Crepas). "Total Quality Metrics: The Revision of an Instrument to Measure the Performance Improvement in Teams" (Jan Streumer, Ton Hendriks, Marije Vermeulen) reports on the design and adjustments of a measurement instrument called Total Quality Metrics that was used to evaluate the performance gains of employees in Polaroid Corporation's division in the Netherlands. A study of the types of evaluation practices being used to evaluate technical training programs in the United States and reasons given by organizations for not evaluating technical training is presented in "Technical Training Program Evaluation: Present Practices in the U.S." (Skip Twitchell, James W. Trott, Jr.). (MN)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 428 237

## 14 Performance Evaluation Process

**Chair: Richard J. Torraco, University of Nebraska**

Assessing the Effectiveness of OJT: A Case Study Approach

*Julie Furst-Bowe, University of Wisconsin-Stout*

*Debra Gates, University of Wisconsin-Stout*

Linking HRD to Agents' Performance in the Insurance Industry

*Teresa M. Palmer, Illinois State University*

*Kenneth J. Crepas, Illinois State University*

Total Quality Metrics: Revision of an Instrument to Measure the Performance Improvement in Teams

*Jan Streumer, University of Twente, Enschede, The Netherlands*

*Ton Hendriks, Polaroid Europe, The Netherlands*

*Marije Vermeulen, KPMG Consulting, The Netherlands*

Technical Training Program Evaluation: Present Practices in the U.S.

*Skip Twitchell, Twitchell Consultants*

*James W. Trott, Jr., Louisiana State University*

AHRD

**ACADEMY OF HUMAN RESOURCE DEVELOPMENT**

**1998 Conference Proceedings**

**March 4 - March 8, 1998**
**Oak Brook, IL**

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY HAS BEEN GRANTED BY
J Furst-Bowe
T Palmer
J Streumer
S Twitchell
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)



ED 428 237



BEST COPY AVAILABLE

# Assessing the Effectiveness of OJT: A Case Study Approach

*Julie Furst-Bowe*
*University of Wisconsin-Stout*

*Debra Gates*
*University of Wisconsin-Stout*

This study used a case study approach to assess the effectiveness of on-the-job training (OJT) in one product line of high tech manufacturing company. A literature review was conducted to identify the characteristics of successful OJT programs. Supervisors, trainers and trainees were interviewed regarding their perceptions of the OJT process. The results indicate there have been many inconsistencies in the implementation of OJT in this product line. Recommendations were made to incorporate greater structure in the planning, delivery and evaluation of OJT. It was also recommended that the organization explore interventions other than OJT programs for skill development.

The economy and changing marketplace have placed numerous demands on manufacturers. These demands have caused many manufacturers to increase flexibility in production and product delivery, expand use of advanced production technologies and increase responsiveness to customer needs. High performing, successful manufacturers depend on employees who can perform complex, ever changing tasks. The demand for employee expertise and the accompanying need to meet changing expertise requirements have placed greater emphasis on training programs to meet specific job requirements (Jacobs and Jones, 1995). Thus, manufacturers are finding several reasons to revisit their OJT (on-the-job training) programs. Costs, flexibility and immediacy are as relevant as ever in comparing OJT to other training options (Barron, 1997). In many manufacturing organizations, more than 80 percent of an employee's job knowledge and skills will typically be learned through OJT (Carnevale and Gainer, 1989). However, despite the widespread use of OJT in the workplace, there have been relatively few efforts to evaluate its effectiveness in light of changing performance requirements.

Company XYZ, a midwest manufacturer of supercomputers, is similar to most manufacturing companies in its use of OJT to develop employee expertise. In the past, this company has tried a variety of OJT approaches to train production workers, from informal activities to structured OJT programs. The late 1980s and the early 1990s proved to be difficult times for Company XYZ, both financially and emotionally. Layoffs, downsizing and budget cuts were all factors which directly affected major organizational functions, including training. During this time, training departments decreased in size and upper level support for training initiatives varied as management teams evolved. There was a need to evaluate current training activities, including OJT, to determine the value of these programs in developing skilled employees.

## Problem Statement

Very little organized research has addressed the topic of OJT (Rothwell & Kazanas, 1994). Few studies have been conducted to evaluate the effectiveness of OJT programs in manufacturing industries. There are many reasons for the lack of interest in this type of research, including the emergence of technology-based training and the notion that OJT is often difficult to distinguish

*Copyright ©1998 Julie A. Furst-Bowe and Debra Gates*

from actual work (Zemke, 1985). The purpose of this study was to evaluate OJT in the manufacturing division of Company XYZ. The results of this evaluation may be used to determine the appropriateness of the use of OJT and recommend improvements to current OJT processes.

## Research Questions

This study sought to answer the following questions:
1.) What are the key characteristics of effective OJT programs as identified by the literature?
2.) What is the current state of OJT at Company XYZ as perceived by employees?
3.) How effective are OJT programs and activities in Company XYZ?
4.) What recommendations can be proposed to improve OJT in Company XYZ?

## Methodology

This study was conducted using case study research methods. Case studies are one of the most prevalent formats for conducting HRD research (Marsick & Watkins, 1997). The case study method may be utilized to investigate a contemporary phenomenon within its real-life context and to describe the unique experience of employees in a given organization (Marsick & Watkins, 1997). In this study, the case study method was used to develop an assessment process for OJT.

The researcher used the following procedures to conduct the case study (Stake, 1994). First, research questions and assumptions were identified. Next, a literature review was conducted to identify the key characteristics associated with effective OJT programs (Jacobs and McGiffin, 1987). The information obtained in the literature review was used to create the interview guides. Separate guides were created for each group interviewed: supervisors, trainers and trainees. A sample was selected and structured interviews were conducted to collect primary data for the study.

A total of 24 employees participated in the interviews: 8 supervisors, 8 trainers and 8 trainees. Each of these employees worked on one of the most mature production lines in the company. Employees from this product line were chosen because the majority of OJT activities center around production workers and because this product line had a long history of OJT experiences. The supervisors that support the four functional areas in the product line were identified. Each supervisor was then asked to identify someone for performed OJT (a trainer) and someone who had received OJT (a learner) within the past six months.

Interview guides were created for each group. Each interview guide was pilot tested with production supervisors and manufacturing trainers who were not participating in the study. Feedback and suggestions for improvement were incorporated into the final instruments. Supervisors were asked about their perceptions and experiences in the following areas: knowledge of policies and procedures regarding OJT, supervisor's role in OJT, skills needed to perform OJT, selection of OJT trainers, planning for OJT, documentation of OJT, evaluation of OJT and rewards for participation in OJT and other training initiatives. Trainers and learners were asked about the same topics. All interviewees were also asked to provide any additional comments or concerns regarding OJT in their work area.

Twenty-four individual interviews were conducted from June 23 to June 27, 1997. Each interview lasted approximately one hour. The interviewer took notes and audiotaped each interview. After all of the interviews were completed, the responses were transcribed, sorted and compiled. Each group of respondents was analyzed separately. In addition, the data were analyzed for emerging patterns and trends across the three groups (Stake, 1994). Conclusions and recommendations were formulated based on the analysis and interpretations of the data collected. As suggested in the procedures for conducting a case study outlined in the *Human Resource Development Research Handbook*, as the data were analyzed and the themes emerged, the

researchers went back to the literature for deeper understanding of the themes and to refine future directions (Marsick and Watkins, 1997).

**Interview Results**

Although there are numerous variations of OJT models, nearly all of the structured models examined in the literature contained the following key characteristics: documented OJT procedures and policies; a management structure with clear roles for supervisors and trainers; work, worker and workplace analysis; training plans; preparation for trainers; designated tools and materials; presentation formats; and evaluation procedures (Jacobs & Jones, 1995; Yang & McLean, 1996). Interview questions were designed around these key characteristics.

Common themes and responses appeared within each of the three groups of interviewees. The following section discusses the common themes surrounding several key characteristics, including documented OJT procedures, roles and skills of the interviewees, trainer selection and preparation, planning for OJT, tools and processes used in OJT, and evaluation procedures.

None of the employees interviewed identified any type of current written documentation concerning OJT policies and procedures. Thus, this product line lacks one of the key elements of an effective OJT training program. However, two respondents mentioned that job descriptions, written training and development plans and other types of documentation used to be available on earlier product lines.

The majority of respondents could clearly describe their perceived roles in the OJT process. In general, supervisors saw themselves as overseeing, monitoring and being ultimately responsible for OJT. However, some supervisors reported little or no involvement in this process. Based on these varying responses, it may be concluded that confusion exists around the role of the supervisor in OJT. The trainers described their role as training employees on how to perform their job efficiently and correctly. The learners believed their roles consisted of learning new tasks, asking questions and attempting to understand the information being presented.

All respondents agreed that there is no formal trainer selection process. Most trainers were selected by supervisors based on their amount of job experience, although both supervisors and trainers indicated that communication skills were the most important skills needed for training. Other skills identified were job knowledge, patience and the ability to simplify tasks. Therefore, it may be concluded that while supervisors know what trainer skills are the most important, they do not select trainers based on these attributes.

Planning for OJT also appears to be a struggle for many supervisors and trainers. Almost all supervisors identify conflicting goals between training and production. Many respondents described a constant tension between training employees and meeting production schedules. Comments such as "there is no plan," "we used to have a plan, but it lacked buy in and support" and "the production schedule always comes first, not training needs" were typical among both supervisors and trainers. Many employees believed that it would be difficult to structure OJT around the new product line due to the number of variations.

The use of processes, tools and materials for OJT varied greatly among respondents. The relationship between processes and tools appears to be direct. Respondents who described their use of training processes also discussed their use of tools and materials. Some trainers were using outlines, job aids, checklists and training matrices to monitor and track trainee progress. Other trainers did not use any processes, tools or materials for OJT.

One reason for this inconsistency may be that nearly all of the trainers developed their training skills on the job. These trainers may not have knowledge of formal training processes and tools. Most trainers said they had learned their roles through experience. Only one respondent had completed a formal trainer development class. However, several trainers mentioned that various tools and training materials had been available on earlier product lines.

There was wide variety in the trainee evaluation practices being utilized. Logbook entries, error recording, shift feedback, observation, questioning and inspections were all identified by supervisors and trainers as ways of obtaining feedback on learner skills. Although all of the

supervisors and trainers discussed the importance of monitoring trainee progress, none of the evaluation techniques were used consistently. Few supervisors described a need to evaluate their trainers. Both trainers and trainees believed that their efforts to deliver or receive training would be tied to their performance reviews. However, supervisors had differing interpretations of how to compensate workers for these activities.

## Discussion

The results of the study indicated that many inconsistencies exist in planning, preparing, delivering and evaluating OJT in the manufacturing division of the Company XYZ. Because this study involved only one organization, it has very limited generalizability. However, the results of this study are not different from the results of other similar studies which have attempted to evaluate the effectiveness of OJT. Several studies have found that in many organizations OJT lacks structure and does not operate within the context of the organizational work system (Jacobs and Jones, 1995).

Specific recommendations to provide greater structure to the OJT program at Company XYZ may include establishing written OJT procedures and policies, providing training for supervisors to clarify their role in OJT, developing a trainer selection process, providing formal training for new and existing trainers and creating an evaluation plan for trainers. A model for OJT could be developed that would include examples of OJT tools that would support the various phases of the process. The implementation of these recommendations would improve the effectiveness and consistency of OJT at Company XYZ, at least for a short time following the implementation of the new training, policies and procedures. However, based on past history and feedback from employees, there is reason to believe it would be difficult to sustain a structured OJT program at Company XYZ.

The individuals responsible for OJT at Company XYZ, and other similar organizations, should ensure the structured OJT process does not become especially lengthy or complicated. The more elaborate and specialized the OJT program, the more reasons supervisors and employees may find to keep it at arm's length (Sisson, 1995). As this study and others have indicated, many supervisors cite a lack of time as an obstacle to implementing structured OJT (Jacobs and Jones, 1995). Some organizations have begun to involve more employees in the design of structured OJT programs as input from employees at all levels of the organization may help assure that the programs have realistic goals and procedures (Walter, 1996).

Structured OJT systems have changed very little since the job instructor training (JIT) approach widely practiced in the 1940s (Sisson, 1995). Although it is recognized that these multiple-step programs do produce skilled employees, in many organizations there is insufficient time to implement the programs properly. What is needed today in many organizations are OJT systems that are so well integrated into existing learning and work processes that OJT does not appear to be a "special HRD program" (Sisson, 1995). The most effective OJT program may be one that has lost its own identity and is simply part of the work system (Sisson, 1995).

Over the past 50 years, there has been little research undertaken in the whole area of OJT (Rothwell & Kazanas, 1994). Most of the literature on OJT consists of descriptions of how structured OJT programs were designed, delivered and perceived by employees in various organizations. There is a need for research that goes beyond the descriptive level. OJT must be evaluated rigorously and objectively in terms of its of its contributions to the organization's learning and performance goals (Jacobs and Jones, 1995). Although structured OJT is a powerful training tool, it may not be the most appropriate tool for enhancing performance in some work environments, particularly with jobs that involve high levels of cognitive processing that should be practiced in authentic complex environments (De Jong, 1996).

Since structured OJT was introduced in organizations, there have been many research advances in the areas of adult learning and performance technology. These advances have allowed researchers and practitioners to design new strategies for enhancing learning and develop new

technologies for supporting performance. Yet many organizations continue to rely on traditional training approaches and fail to utilize new systems that integrate learning and work, despite the inability of employees to implement the traditional systems properly in their current work environments. The most innovative companies focus on developing the knowledge and skills of workers by providing learning and performance support, not structured training (Reinhart, 1997). As companies, such as Company XYZ, attempt to evaluate their current HRD processes and programs in light of changing performance requirements, there is a need to critically examine the role of structured OJT in developing skills and enhancing performance. In some cases, the time and effort spent on structuring or restructuring traditional OJT programs may be better spent on developing integrated, sustainable systems for learning and performance support.

## References


Barron, T. (1997). A structured comeback for OJT. *Technical and Skills Training, 8*(3), 14-17.

De Jong, J.A. (1996). Job instruction reconsidered: Usefulness and limitations of the dominant type of structured OJT. In E. Holton (Ed.), *Proceedings of the 1996 Academy of Human Resource Development Conference*, 235-240.

Carnevale, A.P. & Gainer, L.J. (1989). *The learning enterprise*. Alexandria, VA: American Society for Training and Development.

Jacobs, R.L. & Jones, M.J. (1995). *Structured on-the-job training: Unleashing employee expertise in the workplace*. San Francisco, CA: Berrett-Koehler.

Jacobs, R.L. & McGiffin, T.D. (1987). A human performance system using a structured on-the-job training approach. *Performance and Instruction, 25*(7), 8-11.

Marsick, V. & Watkins, K. (1997). Case study research methods: In R. Swanson and E. Holton (Eds.), *Human resource development research handbook*. San Francisco, CA: Berrett-Koehler.

Reinhart, C. (1997). No more sheep dipping. *Training and Development, 51*(3), 46-51.

Rothwell, W.J. & Kazanas, H.C. (1994). *Improving on-the-job training*. San Francisco, CA: Jossey-Bass.

Sisson, G. (1995). [Review of the book *Structured on-the-job training: Unleashing employee expertise in the workplace*]. *Human Resource Development Quarterly, 6*(3), 320-323.

Stake, R.E. (1994). Case studies. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research methods*. Thousand Oaks, CA: Sage.

Walter, D. (1996). A model for team-driven OJT. *Technical and Skills Training, 22*(10), 93-98.

Yang, J.C. & McLean, G.N. (1996). Structured on-the-job training competencies. In E. Holton (Ed.), *Proceedings of the 1996 Academy of Human Resource Development Conference*, 241 -247.

Zemke, R. (1985). In search of a training philosophy. *Training, 22*(10), 93-98.

## Linking HRD to Agents' Performance in the Insurance Industry

Teresa M. Palmer
Illinois State University

Kenneth J. Crepas
Illinois State University

*Two studies provide information on the nature and extent of insurance industry training or property/casualty and life/health agents, including kinds and amounts of training provided, content of training, and methods used. They show that the industry sees training as a way of addressing performance problems, including legal compliance, attracting and retaining good agents, market complexity issues, and market conduct.*

The ability of an insurance company to sell its products is, to a great extent, dependent on the effectiveness of its agents. For the majority of companies, agents are the most important point of contact with individuals and businesses representing the market for the companies' products. What agents do and how they represent the company have an immediate effect on the company's sales and income for the period and a longer term effect on the company's image and its future profitability.

### Related Literature

Very little information has been published about the training and education of agents provided by insurance companies. A recent review of the literature indexed in the *Business Periodicals Index* yielded over 100 articles in the last ten years, but closer examination of those articles indicated that they contained little substantive information about agent training. What the literature did do was to identify a number of potentially problem perfomance areas for agents that may be addressed, at least in part, by training. Some of these issues included compliance with regulations, attracting and maintaining a strong sales force of agents, the increasing complexity of the marketplace, and market conduct.

***Legal Compliance.*** In the United States, insurance companies and their agents must comply with a complex set of regulations imposed by each state and the federal government. State requirements frequently mandate both pre-licensing training and continuing professional education for agents. Many states have increased education and training requirements for intitial and continued licensing of insurance agents. More stringent regulation is often seen as a response to a general lack of public confidence in the industry's representatives. Continuing education requirements are seen as necessary to professionalism and good front-end training could do a lot to prevent the kinds of damage to the industry's reputation done by poorly prepared agents (King, 1990).

***Attracting and Retaining Agents.*** Several studies suggested that the right training could help companies attract and keep good agents (Wellstead, 1992; Dempsey, 1994; McElroy, et al., 1993). Mc Elroy, et al. (1993) found that the probability of remaining in the profession is positively related to job involvement and professional commitment. They suggested that companies and agencies should spend training resources on activities that enhance these attributes, including such things as increased product knowledge, enhanced self-esteem, reducing role ambiguity and mentoring. These suggestions would seem to be consistent with the views of workplace HRD as empowering and leading to greater professionalism as well as greater commitment to the organization (Dirkx, 1997). As agents' compensation is generally derived from commission on sales, it seems logical that if agents and prospective agents perceive that company

*Copyright @ 1998 Teresa M. Palmer and Kenneth J. Crepas*

sponsored training contributes to achieving success in selling products, that companies offering good training opportunities will be attractive to those agents and prospective agents.

***Market Complexity.*** The insurance industry has been experiencing the same increases in concentration and resultant restructuring as other segments of the financial services industry. The array of products sold by the industry is broad, complex, and continually evolving to meet market needs. Agents require investment licenses to sell many kinds of insurance products today. The actual and potential future increases in the sales of annuities have created a need for agent training in both product knowledge and sales techniques for selling annuities (Dempsey, 1991). The growth of long term care insurance (LTC) has produced, in addition to increased revenues, some very negative publicity for the industry. Agents themselves are calling for more training related to LTC products and sales. Over 70% of 2000 agents surveyed said that they wanted sales support and training seminars (Koco, 1991).

***Market Conduct.*** Several articles in *Consumer Reports* were critical of the practices of a sample of insurance agents. They found that the agents didn't know their products or the market, sold too little or too much coverage, and misrepresented aspects of their policies (King, 1991). Several well-publicized frauds involving agents of major insurance companies in recent years have eroded the public's trust in insurance agents. Recent allegations that Metropolitan Life has been misleading Florida policyholders into buying more expensive life insurance underscore the growing market conduct problems related to agents' sales practices (Scism, 1997).

**Research Questions**

Research to address these performance-related issues would appear to address the category "Linking HRD to Strategy" that was among the most highly ranked of current and future research topics identified by Wentling, et al. (1997). This paper addresses two kinds of research issues/questions. The first is to try to provide a more comprehensive picture of the nature and extent of agent training in the insurance industry, including kinds and amounts of training provided, content of training, methods used, and trends in agent training. Secondly, it examines the perceived relationships between HRD activities for agents sponsored by insurance companies and performance problems related to:

1. compliance with laws and regulations,
2. attracting and retaining good agents,
3. market complexity issues related to the new products developed to meet the increasing competition in the marketplace, and
4. market conduct issues related to maintaining and improving the good name of the company and its products.

**Methodology**

This paper presents data collected from a review of literature and from two surveys. The first survey consisted of a random sample of 100 property-casualty companies selected from those listed in *Best's Insurance Reports (Property/Casualty)*. The preliminary survey instrument, including open-ended questions as well as forced-choice items, was reviewed by a panel of insurance experts. The instrument, a cover letter explaining the purpose of the study, and a return envelope were sent to the directors of agency training at each of the companies in the sample. A follow-up mailing was sent about six weeks after the first. A similar process was used in the second survey, sent to a random sample of 200 life-health insurance companies selected from those listed in *Best's Insurance Reports (Life/Health)*. The questionnaire was revised slightly based upon information received from the first survey. The random selection process used in both surveys resulted in the selection of some companies that do not use agents to sell their products. Analysis of data was carried out separately for the two surveys as they represent two different sectors of the industry and because they were completed about two years apart. Data from the two surveys were tallied, with percentages calculated where appropriate. Responses to the open-ended questions were analyzed using qualitative analysis techniques.

**Findings**

Findings from the two surveys are presented below. Property/casualty and life/health findings related to the nature and extent of agent training are presented separately. Findings related to performance problems, somewhat preliminary at this point, represent analysis of the combined sets of data.

***Property/Casualty Findings.*** There were 27 returns from the first survey, but four of them indicated that they did not use agents, leaving 23 usable returns (23%). Not all of those who responded answered all questions on the survey instrument. Seventy percent of the respondents indicated that training was offered for new agents. The majority of respondents (55%) indicated that new agent training was not required in any of the states in which they write business. Sixteen of the respondents (70%) indicated that training was offered for new agents. Of the 16 insurers responding to the question on length of training, 10 (62.5%) indicated that training was two weeks or more in duration. Therefore, 37.5% of responding insurers provided less that two weeks of training. Pre-reading was required by the majority of insurers, but 75% did not require field experience prior to training. Eighteen insurers indicated that policy analysis was their top priority in agent training. Eight insurers indicated that training on programming needs was provided, and only three provided coverage of competitor's products. Companies were almost evenly divided about administering exams in agents' training programs. The majority of insurers used full-time staff dedicated to agent training and a majority (62%) indicated that they were offering more agent training than five years previously.

Methods used in sales and product training are presented in Table 1. The most heavily relied upon methods for both types of training were printed materials, closely followed by lecture/discussion, speakers, self-study materials, and videos. Mock presentations were used more frequently in sales training and computer simulations were more frequently used in product training. Only two firms reported the use of interactive multi-media in training.

**Table 1 - Methods Used in Sales and Product Training**

| | *Sales Training* | | | | *Product Training* | | | |
|---|---|---|---|---|---|---|---|---|
| | P/C[1] | L/H[2] | Comb.[3] | % | P/C | L/H | Comb. | % |
| Printed materials | 18 | 19 | 37 | 88 | 18 | 22 | 40 | 95 |
| Lecture/discussion | 16 | 15 | 31 | 74 | 18 | 17 | 35 | 83 |
| Mock/presentations | 11 | 15 | 26 | 62 | 8 | 7 | 15 | 36 |
| Videos | 13 | 15 | 28 | 67 | 12 | 10 | 22 | 52 |
| Speakers | 14 | 14 | 28 | 67 | 13 | 13 | 26 | 62 |
| Computer simulation | 5 | 5 | 10 | 24 | 9 | 5 | 14 | 33 |
| Interactive multimedia | 2 | 3 | 5 | 12 | 2 | 4 | 6 | 14 |
| Self-Study materials | 12 | 16 | 28 | 67 | 14 | 17 | 31 | 74 |

[1] P/C = Property/Casualty, n = 19
[2] L/H = Life/Health, n = 23
[3] Combined, n = 42

***Life/Health Findings.*** There was a 15% return rate for the second survey, however several of those responses indicated that they did not have agents, reducing the usable return rate to 12%. Again, not all respondents answered all questions on the survey instrument. Only eight respondents (34%) reported that agent training was mandated by a state within which they operated. Eighty-three percent of those responding provided training for new agents. Half required pre-reading before new agent training but only 33% required field experience prior to training. Ninety percent of responding companies provided training for continuing agents, but only

about one-quarter required continuing agents to attend such training. Seventeen companies included policy analysis in their training, twelve included programming needs, and seven included coverage of competitor's products. Fifty-nine percent of those responding did not administer exams in their agent training programs. Fifty-five percent of responding companies were offering more training than five years previously, while 32% reported no change and 14% reported that they were offering less training than five years previously.

Survey results on methods used in sales and product training are presented in Table 1. Results were very similar to those from the property/casualty sector with print materials and lecture discussion being the most frequently used methods, closely followed by self-study materials. The most noticeable differences between methods used for sales and product training were that mock presentations and videos were noticeably more likely to be used in sales training. Computer simulations were equally likely to be used in the two types of training. Use of interactive multimedia in training of both types was lowest of all the methods listed.

***Findings Related to Performance Problems.*** The preliminary findings related to performance problems come primarily from the thematic analysis of the responses to the open-ended questions on the instrument and the several new questions added to the second survey. The first study included only one open-ended question asking for comments on trends in agent training. The second survey added a second question asking respondents to comment on the most significant change in agent training in the company compared to five years previously.

*Legal Compliance.* More than one-third of the states in which respondents operated required training for new agents. Several respondents addressed this issue in their responses to the open-ended questions:

"More resources are now dedicated to training, and compliance is a more significant challenge."

"Compliance is probably the greatest immediate concern."

"Constant monitoring to keep pace with new products, changes in tax laws, and compliance issues."

"Much more training is required by the states prior to an agent obtaining their license."

*Attracting and Retaining Agents.* Several questions related to this problem were added to the second survey. Twenty-one of 23 respondents (91%) indicated that attracting and keeping good agents was a concern for their companies. Sixty-four percent said that training was *very important* to attracting and keeping good agents, while 32% said it was *important* and 5% said that it was not very important. Only 26% of responding companies have mentorship programs for new agents.

*Market Complexity.* On the first survey, 10% of respondents indicated that they were currently offering less training for agents than five years previously, 29% about the same amount, and 62% either a little more or substantially more. Results of the second survey indicated that 14% of respondents were offering less training than five years previously, 32% about the same, and 55% either a little more or substantially more. Responses to the open-ended questions indicated that training was seen as an appropriate response to issues related to market pressure. Comments included:

"The success of a company will be determined by the training that is available, more training programs will be offered."

"More emphasis on advanced markets."

"Greater need but less time."

"Less is required for the markets we're in; however, products are becoming more complex, requiring more sales literature."

"Development of agent niche marketing is also important, as other sales sources evolve (banks, Internet, etc.)."

*Market Conduct.* In response to the open-ended questions on changes in training in the last five years and emerging trends in training, several respondents specifically mentioned that there was an increased need for training to address market conduct issues for agents.

**Discussion**

Although there had been no specific mention of any of the four problem performance areas in either questionnaire or the cover letters accompanying them, all four were mentioned explicitly in the answers to the open-ended questions. This shows that directors of agency training see a linkage between agent training and solving such performance problems.

An additional theme that emerged strongly in the responses to open-ended questions from both studies dealt with the need to develop more effective training. Several comments indicated a need to move away from passive training to achieve more invovement on the part of agents. Several suggested that there should be more use of computer- based training, and/or on-line, just-in-time materials. There were several mentions that the need for agent training had been increasing, but that resources (both financial and personnel) were declining.

The review of literature provided several articles on computer-based interactive multimedia, but very few firms from either study reported that such programs were being used for agent training in their companies. There is currently no information available to explain the low utilization of technology in training within an industry that is generally characterized as being high-tech.

The varied views expressed by companies about what is needed in agent training may reflect differences among the sectors of the industry (property/casualty versus life/health), changes over the two years of data collection, or differences in how they have been affected by such issues as changing regulation, market pressures, and the need to enhance the image of the agent and the industry with consumers. These issues provide further support for the argument that a higher level of professionalism is necessary among agents, and that additional training may be an essential component of that change.

**References**


Dempsey, Kay I. [1994] Instill agents with unconventional wisdom. *Best's Review* (Life/Health ed.). 95 (November): 86.

Dirkx, John M. [1997] To earn or to learn? The meaning of workplace learning in HRD. *AHRD 1997 Conference Proceedings*, Atlanta, GA: 668-674.

Howard, Lisa S. [1994] U. K. insurers' training costs soaring. *National Underwriter* (Life/Health ed.). 98 (October 3): 61.

King, Carole. [1990] Agent ed: too much too late. *National Underwriter* (Life/Health ed.). 94 (June 18): 13+.

King, Carole. [1991] Consumer's Union does hatchet job on LTC agents. *National Underwriter* (Life/Health ed.). 95 (June 17): 13+.

King, Carole. [1994] New York imposes $500,000 fine on Primerica Life subsidiary. *National Underwriter* (Life/Health ed.). 98 (November 7): 1+.

Koco, Linda. [1991] Agents want LTC training, survey says. *National Underwriter* (Life/Health ed.). 95 (January 7): 7+.

McElroy, James C., Morrow, Paula C., Power, Mark L., Iqbal, Zafar. [1993] Commitment and insurance agents' job perceptions, attitudes, and performance. *The Journal of Risk and Insurance*, 363-384.

Scism, Leslie. [1997] Florida regulators investigate Met Life. *The Wall Street Journal* [Nov. 6]: A3+.

Swanson, R. and Holton III, E., eds. [1997] *Human Resource Development Research Handbook*. Berrett-Kohler Publishers, San Francisco.

_____. [1994] Travelers unit agents could lose licenses amid training probe. *The Wall Street Journal*. Eastern Edition (May 4): A4.

Wellstead, William R. [1992] A career track for sales agents. *Training and Development*. 46: 49-52.

_____. [1992] A review of state continuing education requirements. Staff study by professional insurance agents of America (PIA):Alexandria VA.

Wentling, T., Brinkley, E., and Nelson, E. [1997] Current and future topics in human resource development: linking research with practitioner needs. *AHRD 1997 Conference Proceedings.* 601-608.

# Total Quality Metrics: the revision of an instrument to measure the performance improvement in teams

*Jan Streumer*
*University of Twente, Netherlands*

*Ton Hendriks*
*Polaroid Europe, Netherlands*

*Marije Vermeulen*
*KPMG Consulting, Netherlands*

*Teams are expected to provide organizations with benefits relating to the full utilization of the qualities of the team members, shorter lines of communication, more commitment to the organizational goals and an increase of the effectiveness of the organization. Polaroid Europe introduced so called Business Improvement Teams (BITs) to put the basic message of Total Quality Ownership into practice. A measuring instrument was designed and revised to enable the BITs to constantly improve their performance. This article is focused on the design and adjustments of the measuring instrument, called Total Quality Metrics.*

It can be argued that socio-technology forms the basis for the introduction of teams into organizations. Socio-technology is a design theory that originated in the 1950s. A study of British miners carried out at the end of the 1940s found that people who performed their jobs as part of a team functioned better than those who carried out their work alone. This manifested itself notably in higher productivity and lower rates of absence due to illness among the workers. This finding led to the development of the socio-technological design theory (Van Eijnatten, 1996; Tjepkema, 1997), in which the principles of minimal division of labor and maximum self-direction play a major role.

The term 'team' is defined in various ways. Hackman (1990), for example, defines a team as an intact, clearly-delimited system, within an organizational context, with dependence relationships between the participants, who play different roles, with clear job descriptions and collective responsibility.

## Theoretical Background

Katzenbach & Smith (1993) define a team as a small number of people with complementary skills, who are committed to shared objectives and a joint approach, for which they hold each other accountable.

When definitions of the term 'team' are analyzed for similar elements, a number of things become clear.

1. A team works from a previously defined objective towards a common output (Vroemen, 1995; Katzenbach & Smith, 1993; Senge, 1992). Moreover, it has been shown to be important for teams to continue to be involved in the goal-setting process, in order to ensure that the team embraces the objective both collectively and individually (Katzenbach & Smith, 1993).
2. Team leaders should possess a number of basic skills, as shown below (Vroemen, 1995; Katzenbach & Smith, 1993):
   - technical/functional skills, which are directly linked to the content of the work;
   - problem-solving/management skills that enable team members to solve problems effectively, to make decisions and to see to the planning and structuring of the work;
   - communication skills, such as listening, offering criticism, conflict management and discussion techniques.

*Copyright © 1998 Jan N. Streumer*

3. A team can function synergetically. Synergy is important for a team because the team needs to operate as one, so as to achieve "alignment" (Senge, 1992). Synergy is regarded as an important extra benefit to a team because it enables it to exceed the sum of its individual achievements (Mears, 1994).
4. There must be job dependence in a team. A distinction can be made between serial job dependence, where the output of one team member is the input of another, and parallel job dependence, where there is simultaneous effort (Vroemen, 1995). The extent of job dependence too may vary per team: in some teams, members are able to take over each other's jobs almost completely in the event of absence, which indicates a form of job design that is characterized by job rotation, job enlargement and job enrichment (Mears, 1994). In this last case there is a slight measure of job interdependence within the team, given that members are able to take over each other's jobs. However, it is also possible that team members complement each other in the performance of their jobs and that the absence of a team member may not be so easily resolved. In such a case there is extreme interdependence because team members are not able to take over each other's jobs.
5. A team should be of a certain size. Literature on this subject is somewhat ambiguous. As a lower limit, two members are mentioned, and twenty as an upper limit (Katzenbach & Smith, 1993). A minimum number of team members is important in order to be able to achieve job differentiation within a team; a maximum number of team members is advisable to prevent communication problems and the formation of subgroups.

The literature analyzed reveals that teams are expected to provide organizations with benefits relating to the following factors:

1. *Skills*: a team brings together complementary skills and experiences, which by definition exceed those of each individual team member (Katzenbach & Smith, 1993). Teamwork allows the qualities of the team members to be utilized to the full (Van Ewijk-Hoevenaars, Van Jaarsveld & den Hartog, 1995).
2. *Communication*: by developing collective objectives and strategies, a form of communication is created that provides a great deal of support when solving problems and exercising initiative. Within teams both vertical and lateral communication occurs, cross-departmental link-ups take place, more ideas are generated and, lastly, there is mutual respect (Mears, 1994). Teams also lead to a flatter organizational structure, resulting in shorter lines of communication and enabling decisions to be reached more quickly (Osburn, Moran, Musselwhite & Zenger, 1990; Van Ewijk-Hoevenaars et al, 1995).
3. *Flexibility*: there is a close link between clear communication and the flexibility of teams. As far as flexibility is concerned, teams are said to possess those skills, information and motivation which enable them to adapt easily to change, thus allowing the organization to respond flexibly to changes coming from its surroundings (Osburn, Moran, Musselwhite & Zenger, 1990).
4. *Participation:* working in a team produces a positive effect on morale; it creates opportunities and provides a greater degree of job satisfaction (Mears, 1994). There is also a unique social dimension within a team which has a stimulating effect on the economic and management aspects of the work. Within a team there is a great degree of mutual trust, which can break down barriers (Katzenbach & Smith, 1993).
5. *Effectiveness*: working together in a team ensures that better quality solutions can be implemented and that there is 'process ownership', where the team as a whole bears responsibility for the entire manufacturing process. As a result, team members are more motivated to perform better and the effectiveness of the organization will increase (Osburn, Moran, Musselwhite & Zenger, 1990).
6. *Satisfaction*: a team derives more pleasure and satisfaction from its work than does a single individual. Customers will be more quickly satisfied since teams can generally react rapidly and respond to changes, which leads to improved quality being provided (Osburn, Moran, Musselwhite & Zenger, 1990).
7. *Commitment*: as a result of their greater responsibility, team members are more involved in the whole manufacturing process and will show more commitment to general organizational goals.

Den Hartog (In Van Ewijk-Hoevenaars et al, 1995), however, observes that the above benefits can only be achieved when teams operate in an integral vision of organizational renewal. What is

important here, if long-term effects are to be achieved, is that teams are set up not only in accordance with the principles of work structuring, job enlargement, job enrichment and job rotation, but that they are also seen as an integral part of the organization. This involves a shared belief within the whole organization, from top management to shop floor, that teams are essential in order to be able to operate effectively as an organization.

**Problem Statement**

In 1992 Polaroid Corporation launched the Total Quality Ownership Awareness Program to make employees aware of the necessity for change and the role that they could play in this.

In March 1993 Polaroid Enschede (the Netherlands) took over the initiative from America to adapt the organization with the aid of Total Quality Ownership (TQO) to a dynamic environment, which is characterized by changing markets and increasing competition.

*Total Quality ownership is interpreted within Polaroid Enschede (the Netherlands) as:*
*"The way in which we do business, both within the company and with others, so that we are always able to supply products and services which meet the expectations of our customers. These efforts are continually focused on achieving and maintaining the satisfaction of the shareholders, the customers and the employees of our company."* (Polaroid Europa BV, 1996a).

A process of change such as TQO will develop by trial and error and will bring with it uncertainty and resistance. It is therefore important for all employees to be informed of the need for a process of change such as TQO. A so-called 'awareness-phase' is necessary to achieve this. This resulted in all Polaroid employees following a TQO Awareness Program, the aim of which was to make employees aware of the necessity for change and the role that each employee had to play in it.

After the 'awareness phase' had been completed, Polaroid Enschede embarked on the introduction of Business Improvement Teams (BITs) in 1994. A BIT is a team of between three and fifteen employees (the BIT-members) who, using TQO principles and the success factors related to these, supply a product or service to the customer. A BIT may be seen as a work unit within a department, where the employees have one or more work processes in common (for example, within the Assembly or Purchasing Department). By means of regular consultation in so-called BIT-meetings, activities are undertaken to continually improve not only the quality of products and/or services, but also the quality of the work process.

The aim of the BITs is to put the basic message of TQO into practice, as well as to create a consultation and communication structure and culture between different work units. The members of a BIT consult each other in a BIT-meeting, which is led by an immediate superior (the BIT leader) and an external advisor from the Training & Development Department. They work in accordance with a dynamic model consisting of four steps, aimed at both product and process:

*Step 1: Review of team output*
This step is characterized by an evaluation of the concrete result or product. Efforts should be made to find out if the goals set have been met, how the communication process has progressed (with both internal and external customers), what went well and what went wrong and what could be learned from all this.

*Step 2: Review of teamwork*
This step is characterized by an evaluation of the process by which the goals have been achieved. In doing this checks are made to see if there has been sufficient consultation, regular communications, clear objectives, cooperation or opposition, a smooth process, or conversely, individual efforts that have hampered cooperation.

*Step 3: View of teamwork:*
The work consultation (the BIT-meeting) provides an ideal opportunity to find out how people should work together, what role employees should play in this and how and with whom there should be communications. Making agreements about these matters has a positive effect on the cooperation process.

*Step 4: View of team output*
Objectives needs to be revised, with the results already achieved acting as the starting point. What is important is that objectives are clearly defined, that it is laid down how and when they should be achieved and that the allocation of people and resources is indicated.

The measuring instrument TQMetrics was designed to enable BITs to constantly improve within the framework of TQO (Polaroid Europa BV, 1996b).

TQMetrics is a measuring instrument, composed of links and sub-links, which uses eight critical success factors of teamwork to chart the performance of a team.

The critical success factors are:

1. Clear *communication*, aimed at both communication within the team and with other teams and managers.
2. *Cooperation*, both within the team and with other teams. A close partnership- relationship is important.
3. *Team goals*, drawn up in accordance with SMART-criteria (Specific, Measurable, Achievable, Result-oriented, Timebound).
4. A good partnership-relationship with the team's customers by means of a *customer-oriented approach*. This involves a clear and balanced cooperative relationship with the team's customers, both inside and outside the company.
5. *Supplier-oriented approach*, the same applies here as for the customer-oriented approach, but then only for suppliers inside and outside the company.
6. *Leadership* of the team should be even-handed and provide coaching, and a balance should be struck between the participatory, socio-emotional and job-oriented aspects of leadership.
7. *Work processes* that relate to the production of products and services. The design and implementation of these processes by team members have an influence on the team's performance.
8. *Continual improvement of the know-how and skill* of each individual team member, which again provides a greater guarantee of future success. By means of a complex and constantly changing environment, this know-how and skill are continually adapted and it is important that team members are open to new learning experiences.

The answers from the BITs can be input into a database program to generate reports. These reports then form the basis for formulating improvement opportunities, for example, as regards cooperation or communication within teams.

The quality of the existing TQMetrics was established in collaboration with the Training & Development Department of Polaroid Enschede, the Netherlands. TQMetrics was then revised in the sense that: (a) the missing essential indicators of effective team performance were added; (b) the reliability of TQMetrics was improved and (c) the validity and reliability of the revised instrument were established.

***Research Design*** In order to be able to answer the research questions, first and foremost, the reliability and validity of the existing TQMetrics was established.

Next, a literature search was carried out into indicators of effective team performance, which were then compared with the indicators that formed the basis of the existing TQMetrics. The final results of these two activities have led to adjustments being made to the existing TQMetrics. These adjustments resulted in an updated TQMetrics, which was then evaluated with the aid of a sample survey of three BIT-teams.

The sample survey was made up of the three selected BIT-teams, since the remaining teams in the organization had only recently used the old version of TQMetrics. For reasons of validity, it was decided to exclude these teams from participation.

Then, using the SPSS statistical analysis program, reliability and factor analyses were carried out on the TQMetrics scales and sub-scales.

Finally, conclusions were reached on the basis of these statistical analyses and recommendations were then made to further improve the revised TQMetrics.

***Research Findings***

*Reliability and Validation of TQMetrics* A good measuring instrument is both reliable and valid (Spector, 1992). Reliability ensures that the instrument consistently measures a particular construct, but it does not ensure that it measures what it is supposed to measure. This last aspect relates to the term validity.

The reliability analyses of the TQMetrics scales and sub-scales were carried out with Cronbach's alpha technique from the Statistical Package for Social Sciences.

TQMetrics consists of eight scales, which correspond to the critical success factors for TQO mentioned above. Table 1 reflects the values of the scales that were calculated using Cronbach's alpha, the number of sub-scales and the number of items per scale.

**Table 1. Survey of reliability analysis per questionnaire**

| *Scales* | *Cronbach's alpha* | *Number of scales* | *# Items* |
|---|---|---|---|
| 1. Communication | .87 | 4 | 21 |
| 2. Cooperation | .89 | 6 | 20 |
| 3. Team objectives | .93 | 4 | 18 |
| 4. (Internal) customers | .94 | 5 | 20 |
| 5. (Internal) suppliers | .95 | 5 | 19 |
| 6. Work processes | .93 | 3 | 20 |
| 7. Leadership | .98 | 3 | 20 |
| 8. Continual improvement | .93 | 3 | 19 |

From the above data the average alpha proves to be .93. On the basis of the reliability analysis, it may be concluded that the scales are reliable (.70 is retained as the lower limit).

As far as the analysis of the sub-scales is concerned, it is notable that all the sub-scales have reasonable to high alphas (all above .70!), with the exception of the 'know your customers' sub-scale (from the 'Suppliers' scale); the 'respect and appreciation', 'meeting commitments' and 'accountability' sub-scales (from the 'Cooperation' scale); the 'measuring results' and 'communicating results' sub-scales (from the 'Team objectives' scale); and, finally, the 'know the customer's work process' sub-scale (from the 'Work processes' scale).

It was also established that a limited number of sub-scales contain too few items, which might partly explain the alphas that are too low.

During the validation of the existing TQMetrics, checks were made using factor analysis as to whether the eight scales are able to measure properly the underlying constructs.

To gain a clearer overview of the number of factors and the extent to which the variables (the various questions) load on this, a rotated factor matrix was carried out on the selected number of factors (Norusis,1993; Kinnear & Greay, 1994). The standard method that is used during rotation is the varimax method, which tries as much as possible to minimize the number of variables that load high on a factor. As a result, the factors can be better interpreted.

Two sorts of factor analysis were carried out. During the first, all the questions from TQMetrics were input one after the other, with the aim of finding out the number of underlying factors (scales) within TQMetrics. The second factor analysis consisted of analyses per existing scale, with the aim of discovering the number of underlying sub-scales within the TQMetrics scales.

The factor analysis that was carried out reveals that nine factors can be distinguished instead of the eight scales that occur in the existing TQMetrics. Table 2 shows three factors and the underlying constructs.

**Table 2. Survey of definitive underlying factors**

| *Factors* | *Underlying constructs* |
|---|---|
| Factor 1 | Leadership |
| Factor 2 | Suppliers |
| Factor 3 | Communication |
| Factor 4 | Customers |
| Factor 5 | Objectives |
| Factor 6 | Continual improvement |
| Factor 7 | Work processes/Tailoring to meet the demands and wishes of customers |
| Factor 8 | Team's method of working |
| Factor 9 | Active attitude |

The data show that the questions from the original 'Communication' and 'Cooperation' scales all load on factor 3 (Communication) and that there is thus no separate 'Cooperation' factor. A second observation relates to the addition of a ninth factor, concerning the employees making an active contribution to team performance.

To sum up, with regard to the number of underlying sub-scales per scale, the analysis results indicate that the number of existing sub-scales per scale (in the present TQMetrics) is generally too high.

*Indicators of Effective Team Performance* A thorough literature search, focused on the concept of team performance, resulted in a new list of 21 indicators of effective team performance being drawn up. This relates to the following indicators: 1) size of team, 2) involvement of team members; 3) internal team communication; 4) organizational culture; 5) explicit job description; 6) clarity as regards desired behavior; 7) conflict management; 8) leadership; 9) basic skills; 10) procedures/methods of working; 11) shared aims; 12) joint responsibility; 13) flexibility; 14) delegation of competencies; 15) (on-the-job) training; 16) support; 17) achievement standards; 18) feedback; 19) cooperation with other teams; 20) remuneration structure; 21) logical job design.

This list of indicators, together with the final results of the reliability analyses and the validation of the scales and sub-scales, will serve as input for revising the existing TQMetrics.

*The Revision of TQMetrics* A comparison of the success factors used during the development of the present TQMetrics with the list of indicators of effective team performance drawn up on the basis of a literature search has resulted in a list of supplementary substantive demands which the revised TQMetrics must meet. These are demands which required attention per scale category, mainly in the form of the transfer of existing items to other (sub)scales and the addition of new items about missing (aspects of) indicators.

On the basis of the reliability and factor analysis carried out of the scales and sub-scales of the existing TQMetrics, recommendations were made in order to achieve a new scale and sub-scale ranking; the number of scales remained as it was at eight. The demands and recommendations were discussed with two employees from Polaroid Enschede's Training & Development Department and, finally, a new TQMetrics prototype was drawn up.

*Evaluation of the New TQMetrics Prototype* The TQMetrics prototype was sent to the three selected teams (from the same department), with a total of 45 employees. The data from the questionnaires that they returned were then analyzed for reliability with Cronbach's alpha (from the SPSS-offer). This involved two analyses being carried out: one on the scales and one on the sub-scales.

The results of the reliability analysis on scale level are shown in Table 3.

**Table 3. Survey of reliability analysis scales updated version of TQMetrics**

| Scales | Change in Cronbach's alpha | | | Change in items |
|---|---|---|---|---|
| | old alpha | | new alpha | |
| 1. Communication | .87 | → | .90 | 21 → 18 |
| 2. Cooperation | .89 | → | .89 | 20 → 21 |
| 3. Team objectives | .93 | → | .93 | 18 → 16 |
| 4. Internal customers | .94 | → | .93 | 20 → 16 |
| 5. Internal suppliers | .95 | → | .91 | 19 → 12 |
| 6. Work processes | .93 | → | .87 | 20 → 20 |
| 7. Leadership | .98 | → | .95 | 20 → 19 |
| 8. Continual improvement | .93 | → | .97 | 19 → 14 |

Table 3 also shows the old and new alphas and the changes that have occurred in the number of items.

As regards the reliability analysis of the scales, it can be said that reliability is still very high, compared with that of the old TQMetrics, which indicates that the reliability of the scales is good. The average alpha of the TQMetrics scales was .93, the average alpha of the new TQMetrics scales is .92. On average, thus, the alpha has decreased slightly. It should, however, be noted here that this may be linked to the fact that the reliability analysis of the new TQMetrics was carried out on a selection of three teams from one department with a total of 45 test subjects, whereas the reliability analysis of the old TQMetrics was carried out on a total of 352 test subjects from all departments. The difference in the size and representation of the sample survey may serve as an explanation for the (slight) drop in the alpha of the new TQMetrics sub-scales.

The results of the reliability analysis at sub-scale level have been incorporated into Table 4.

The results of the reliability analysis of the sub-scales of the new TQMetrics compared with those of the old TQMetrics are more striking. In general, it can be said that the sub-scales have somewhat decreased in reliability. An exception to this are the eight sub-scales that have increased in reliability and the two sub-scales that have decreased in reliability. As already mentioned above, one explanation which may be given for this decrease in reliability is the fact that the new TQMetrics was only tested on 45 subjects, compared with 352 in the old TQMetrics. An additional factor is that many sub-scales decreased in the number of items, which may also explain the decrease in the reliability of a number of sub-scales, since a longer test (with more items) is generally more reliable (Allen &Yen, 1979). Furthermore, it appears from a closer analysis of the scoring behavior of the 45 test subjects that about half may well have answered in a socially desirable manner. This is caused mainly by one specific team that is perceived as "very law-abiding". This too may serve as a possible explanation for the decrease in the reliability of the sub-scales.

***Conclusions and Recommendations*** In this present study efforts were made to discover whether the existing TQMetrics met reliability and validity requirements and whether TQMetrics was based on indicators of a sufficiently high quality.

On the basis of a literature search TQMetrics was deemed not to include all the aspects of effective team performance. Despite the fact that the scales in TQMetrics had a high degree of reliability, the reliability of the sub-scales could be considerably improved. The same applied to the validity of TQMetrics. The validity of TQMetrics was reasonable, but could be substantially improved by a change in the composition of the sub-scales.

The revised TQMetrics was then evaluated, from which it could be concluded that the updated version contained the correct indicators of effective team performance and proved to be a reliable measuring instrument. The fact that a number of scales and sub-scales have a lower level of reliability than in the old version of TQMetrics may be attributed to the sample survey of three teams. The number of respondents involved in this sample survey is much smaller than the

number of respondents on which the reliability of the old version of TQMetrics was based. Secondly, a large number of sub-scales in the updated version of TQMetrics consist of fewer items than was previously the case. Thirdly, the socially desirable answering behavior of the respondents may well have played a role. A fourth factor is the standardized nature of the work process of the evaluation groups, which is partly at variance with the basic principles of team work. Finally, it can again be observed that the sample survey is not representative of the total number of teams within Polaroid Enschede.

The above statements may be regarded as possible causes of the decrease in the reliability within several scales and sub-scales of the new TQMetrics compared with the old version. It is therefore to be recommended, in due course, when a representative number of employees and teams have filled out the updated TQMetrics, that the reliability and validity of the measuring instrument are again studied in order to be able to reach valid conclusions.

**Table 4. Survey of reliability analyses sub-scales updated version of TQMetrics, in comparison to old reliability**

| Sub-scales | Changes in Cronbach's alpha | | | Changes in items |
|---|---|---|---|---|
| | old alpha | | new alpha | |
| Communication | | | | |
| • Horizontal communication | .78 | → | .81 | 10 → 8 |
| • Team consultation | .88 | → | .76 | 17 → 5 |
| • Openness in the team | .75 | → | .79 | 7 → 6 |
| Cooperation | | | | |
| • Cooperation in team | .85 | → | .78 | 11 → 6 |
| • Cooperation with other teams | .89 | → | .80 | 13 → 4 |
| • Responsibility | .63 | → | .66 | 4 → 4 |
| • Respect and appreciation | .40 | → | .54 | 3 → 6 |
| Team objectives | | | | |
| • Formulate objectives | .85 | → | .84 | 6 → 4 |
| • Communicate objectives | .77 | → | .79 | 3 → 4 |
| • Operationalize objectives | | .85 (new) | | 4 |
| Internal customers | | | | |
| • Know your customers | .73 | → | .88 | 5 → 9 |
| • Meet expectations | .88 | → | .88 | 3 → 4 |
| Internal suppliers | | | | |
| • Know your suppliers | .63 | → | .84 | 5 → 9 |
| • Communicate expectations | .77 | → | .86 | 9 → 8 |
| Work activities | | | | |
| • Know your work process | .84 | → | .41 | 3 → 5 |
| • Improve work processes | | .85 (new) | | 7 |
| Leadership | | | | |
| • Participatory leadership | .86 | → | .84 | 5 → 5 |
| • Soco-emotional leadership | .97 | → | .93 | 11 → 10 |
| • Job-oriented leadership | .94 | → | .82 | 4 → 6 |
| Continual improvement | | | | |
| • Willingness to change | .82 | → | .43 | 9 → 4 |
| • Open attitude to learning | .90 | → | .91 | 11 → 7 |
| • Team learning | .82 | → | .77 | 2 → 4 |
| • Active attitude | | .66 (new) | | 5 |

**References**

Allen, M.J. & W.M. Yen (1979). *Introduction to measurement theory*. Monterey, California: Brooks/Cole Pub. Co.

Eijnatten, F.M. van (1996) *Sociotechnisch ontwerpen* [Socio-technological designing]. Utrecht: Uitgeverij Lemma BV.

Van Ewijk-Hoevenaars, A.M., Jaarsveld, J.C.M. van & J.F. den Hertog (1995). *Naar Eenvoud in Organisatie: Werken met zelfsturende eenheden.* [Towards Simplicity in Organization: Working with self-directed units]. Deventer: Kluwer Bedrijfswetenschappen.

Hackman, J.R. (Ed.) (199). *Groups that work (and Those That Don't): Creating Conditions for Effective Teamwork.* San Francisco: Jossey-Bass.

Katzenbach, J.R. & D.K. Smith (1993). *Het geheim van teams. Een bedrijf van wereldklasse creëren.* [The secret of teams. Creating a world-class company]. McKinsey & Company, Inc.

Kinnear, P.R. & C.D. Gray (1994). *SPSS for Windows made simple.* East Sussex: Psychology Press Publishers.

Mears, P. (1994). *Organization Teams. Building Continuous Quality Improvement.* Florida: St. Lucie Press.

Norusis, M.J. (1993). *SPSS for Windows: Professional Statistics, Release 6.0.* Chicago: SPSS Inc.

Orsburn, J.D., Moran, L., Musselwhite, E. & J.H. Zenger (1990). *Self-directed work teams: the new American challenge.* Illinois: Business One Irwin.

Polaroid (Europa) B.V. (1996a). *Total Quality Ownership Awareness Training. Introductie tot Total Quality Ownership. Deelnemers Werkboek.* [Total Quality Ownership Awareness Training. Introduction to Total Quality Ownership. Participants' Workbook. Enschede: Polaroid (Europa) B.V., Training & Development Department.

Polaroid (Europa) B.V. (1996b). *Total Quality Metric. Het eigen teamwerk gemeten.* [Measuring one's own teamwork]. Handleiding TQMetrics. Enschede: Polaroid (Europa) B.V., Training & Development Department.

Senge, P.M. (1992). *De Vijfde Discipline; de kunst en praktijk van de lerende organisatie.* [The Fifth Discipline: the art and practice of the learning organization]. Schiedam: Scriptum Books.

Spector, P.E. (1992). *Summated Rating Scale Construction: an Introduction.* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-082). Newbury park, CA: Sage Publications, Inc.

Tjepkema, S. (1997). Zelfsturende teams: naar synergy tussen leren en werken? [Self-directed teams: towards synergy between learning and work?]. In: *Opleiding & Ontwikkeling, April 1997.*

Vroemen, M.W. (1995). *Werken in teams.* [Working in teams]. Deventer: Kluwer Bedrijfswetenschappen.

# Technical Training Program Evaluation: Present Practices in the U.S.

*Skip Twitchell*
*Twitchell Consultants*

*James W. Trott, Jr.*
*Louisiana State University*

*This study sought to determine the types of evaluation practices being employed to evaluate technical training programs in the United States and the degree to which various evaluation practices were being utilized. Technical trainers were surveyed because technical training is often believed to be easier to evaluate and no prior research was available. Data were gathered on the types and methods of evaluation used, reasons for not evaluating, organizational training practices and selected demographic information. The survey findings supported the lack of training evaluation reported in other more broad studies. These findings suggest a new research agenda on organizational training decision processes is needed.*

It is now well established in HRD research that many training programs are not evaluated. It has become a common lament of HRD researchers that evaluation is essential for effective HRD, but not often conducted. One barrier often mentioned is that evaluation is too difficult because the outcomes of training are hard to identify. For example, management training and soft skills training have less clear outcomes. On the other hand, evaluating training for those whose work produces objectively measurable outcomes, such as technical or clerical employees, lends itself to quantitative evaluation. Technical workers produce goods, construct physical works (public or private), repair machines, operate equipment, and develop computer programs. All of this work produces objectively measurable outcomes. The ability to do each of these jobs can be evaluated using objective outcomes.

Because technical training is designed to produce objectively measurable changes in skills and knowledge, it would be logical to think that evaluation would be more commonly conducted on technical training courses if, in fact, the difficulty of identifying training outcomes is a key barrier to evaluation in practice. Unfortunately, no prior research has established whether technical trainers' evaluation practices are similar or different.

The purpose of this study was to examine evaluation practices of technical trainers. More specifically, the study addressed the following research questions:

1. To what extent are each of the four levels of evaluation, as defined by Kirkpatrick(1994), used to evaluate technical training in business and industry?
2. What impediments exit to the use of evaluation at all levels?
3. What methods are used at levels 2,3,and 4 ?
4. Is there a relationship between selected demographics, training practices, and the value placed on training by training managers, and what levels of evaluation are used?

## Related Literature

The use of evaluation in business and industry represents a subset of the broader field of evaluation literature with a greater emphasis on summative measures rather than formative measures. Several authors have commented on the use of evaluation by business and industry. The focus of most evaluations in business and industry is on measuring a program's effect on (a) the participants,

©copyright, S. Twitchell & J. Trott, 1998

(b) the participant's work, and (c) the organization (Brinkerhoff, 1991; Broad & Newstrom, 1992; Dixon, 1990; Kirkpatrick, 1994; Phillips, 1991).

"It's probably fair to say that the bulk of all employee training programs conducted in the United States are evaluated only at level 1, if at all. Of the rest, the majority measured only at level 2" (Gordon, 1991, p.21) Very little training evaluation is accomplished across American industry. (Dixon, 1987; Phillips, 1991; Robinson & Robinson, 1989). Dixon (1987) observes that when evaluation is done, the participant reaction form is the most frequently used vehicle. This observation has been confirmed by others.

Robinson and Robinson (1989) reported that in a survey of 150 training directors, 77% reported using level 1 evaluation in 81 to 100% of their courses. Ninety-seven percent reported using level 1 evaluation to some extent in their training courses. In technical training, use of level 1 evaluation may be even lower. The Corporate HRD Executive Survey of the American Society of Training and Development (1989), reported that in technical training only 57% used participant reaction forms. This survey reports a much lower use of participant feedback in technical skills training than in management training where the rate of participant feedback evaluation is 92%.

Robinson & Robinson (1989) report that only 36% of the training directors surveyed used level 2 evaluation in 60% or more of their courses. The ASTD Corporate HRD Executive Survey (1989) reports that for technical skills training, 45% of the companies used this form of evaluation. Survey Report #11 reflects a greater amount of evaluation at level 2 than do other surveys. This is not surprising in that technical training often lends itself to the use of objective measures. Likewise, the literature is replete with references on developing test of various types at this level.

Carnevele & Schulz (1990) place the use of level 3 evaluation at between 10 and 15% based on two different surveys. Robinson & Robinson (1989) report that six and 13% of training managers use level 3 evaluation for 80% of the courses taught. In the same report, 20% of courses were evaluated at level 3 by 31% of training managers. The ASTD'S Corporate HRD Executive Survey (1989) reported evaluation methods for level 3 were post course tests, performance records, and supervisor feedback reported by 37, 31, and 29 percent of the companies responding respectively.

Of the surveys found in the literature, only two reports discussed evaluation at level 4. Carnevale and Schulz (1990) report that "Employee training was only evaluated at the organizational result level about 25% of the time, despite new pressures on training practitioners to assess the economic worth of HRD activities." Robinson and Robinson (1989) report that for 80% of the courses taught, less than 6% of the training managers use level 4 evaluation and 59% of managers report never using level 4 evaluation. Most discussions of cost-benefit analysis or return on investment deal more with cost accounting for training than they do on estimating the dollar benefit of training.

## Methodology

***Survey Population and Sample*** The population for this study consisted of all U.S. business and industry providing technical training. The accessible population was the 2,569 U.S. business and industry organizations with ASTD members who were in the Technical and Skills Training professional practice area. To identify this population, all consultants, schools, and government organizations were removed from the ASTD mailing list for the Technical and Skills Training professional practice area. Schools and government organizations were removed because they are not members of business or industry. Consultants were removed because they are both businesses that train their own personnel and the personnel from other organizations. It was felt that they would provide a mixture of data from the two groups and would not be representative of business and industry. Population errors included some consultants among those surveyed because the business names and person's title did not provide information about the business' function. The data from these surveys were not used, but notes returned with the surveys supported the assumption that consultants were a special case and that the data collected from this group would include a mixture of internal and external training data.

When multiple members existed within an organization, the survey was sent to a single member. Selection was in the following order; (1) the member listed as managing or directing the technical training function, (2) highest position in the training function, (3) the position deemed most likely to be connected to technical training. For identical titles, the choice was made using a table of random numbers. Those with no title were used only if no other alternative was available. The organizations with no member's name listed were excluded from the sample, since there was no way to assure that each mailing went to the same person.

This accessible population was chosen because it should represent organizations in which those persons responsible for the technical training function are well informed as to the current practices in industry. As such, the population should report higher uses of evaluation than the total population. As low levels of use were expected based on the literature, this scheme provided a best case scenario of the use of evaluation in the technical training area.

A random sample was drawn from the accessible population. Sample size was 334 after using Cochran's formula for small populations and providing for a $p = .05$. Since this was expected to be a reluctant population, the entire population was randomized which allowed for replacement of those unwilling to participate and for frame errors. In a similar study, Gutek's (1988) original response rate was 18 % and after phone contact with all the non-respondents the final sample consisted of 46 % of the drawn sample. Schilling's (1991) original response rate from an unsolicited set of mailed surveys was 3 %. In that study, a non-random sample was used to increase the number of respondents.

***Instrumentation and Data Collection*** The data for the study were collected by survey. The instrument was developed for the study and revised based on input from a review panel with representatives from university researchers, training managers, training specialists, and training experts from business and industry. The survey consisted of seven sections. Sections 1 through 4 included questions on the percent of programs evaluated at each Level, types of evaluations used, and reasons for not evaluating at each Level. Despite criticism of the four level model (Holton, 1996), it was used as an organizing framework because it is the most widely known model and technical trainers would be most familiar with it. However, explanatory text and examples were also included in case respondents were not familiar with the four level model.

Section five consisted of questions used to determine the training practices of the organization. Section six gathered information on the respondents' thoughts on the value of evaluation. Section seven gathered demographic data. Questions in sections five and six were used to gather information on evaluation practices that were not a part of Kirkpatrick's model.

The data collection procedure used was based on Dillman's (1978) TDM (Total Design Method) for mailed surveys. The survey was followed by a mailing of reminder cards to the entire sample one week after the initial survey was mailed. Three weeks after the original mailing, a second mailing of surveys to the non-respondents was sent with a new cover letter explaining the importance of their reply to the quality of the results of this study. Seven weeks after the first mailing, a second follow-up letter was sent to all non-respondents.

A random sample of 20 non-respondents was contacted by phone and asked to respond to a sample of 22 questions drawn from the original questionnaires. The questions used were selected based on their importance to the objectives of the study. If a non-respondent could not be contacted or refused to participate they were replaced.

## Findings

A total of 146 surveys were returned for an overall response part of 42%. When adjusted for unusable or incomplete returns, a usable response rate of 35% was realized. Non-Respondents were found to differ significantly from respondents on only two of the 22 items (a) the percentage of programs in which evaluation was the first step in program development and (b) the percentage of programs dependent on evaluation for continued funding. Of the two, the most significant related to funding. Four percent of the respondents' programs were dependent on evaluation for continued funding while 35% of the non-respondent's programs were reported as depending on evaluation for continued

funding.

***Selected Demographics*** Responding organizations ranged in size from less than 50 employees to 50,000. Organizations ranging in size from greater than 100 employees to less than 5,000 accounted for 69% of the respondents.

Respondents were asked to indicate the industry group that best described the organization's function. Five function groups were each selected by more than five percent or more of the respondents. These were automotive, banking, chemicals and pharmaceuticals, computer-data processing, manufacturing, and utilities.

***Research Question 1 - To what extent are each of the four levels of evaluation, as defined by Kirkpatrick (1994), used to evaluate technical training in business and industry?*** Two approaches are used to answer this question. First, the data is examined by percent of programs employing each level of evaluation. Then, the data are reanalyzed by percentage of organizations using each level to some extent.

*Percent of Programs Evaluated* For each of the four levels of evaluation, each respondent was asked to indicate the percentage of their training programs that employed each level of evaluation. This reporting scheme also permitted the classification of respondents based on their use, to any degree, of a particular level of evaluation. The mean percentage of programs indicating the use of level of evaluation to some extent is shown below and in Table .

| Percent of Programs Evaluated at Each Level | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| | 72.7% | 47.1% | 33.7% | 20.8% |

The declining degree of use from level 1 through level 4 is consistent with the literature thus suggesting that use practices in technical training do not differ greatly from that reported in the literature for training as a whole.

**Table 1 - Frequencies and Percentages of Programs Using Four Levels of Evaluation for Selected Percentage Intervals**

| Intervals | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| 0% | 8.1% | 16.2% | 35.1% | 46.8% |
| 1-19% | 7.2% | 15.3% | 14.4% | 21.6% |
| 20-39% | 2.7% | 11.7% | 13.5% | 9.9% |
| 40-59% | 9.9% | 14.4% | 16.2% | 4.5% |
| 60-79% | 8.1% | 13.5% | 5.4% | 6.3% |
| 80-100% | 64.0% | 28.8% | 15.3% | 10.8% |

*Percent of Organizations Using Each Level* When the data from percentage of programs using an evaluation Level is reported as the percentage of organizations using an evaluation Level, the definition of use greatly affects the outcome. The data collected in this study were separated into three use ranges. This was done by choosing three different minimum percentages of programs required to qualify a responding organization as using that level of evaluation. The organizations were separated into ranges using the following criteria:

1. organizations using the Level in any percent of their programs other than zero (some use)
2. organizations using the Level in 60 % or more of their programs (more than 60%)
3. organizations using the Level in 80 % or more of their programs (most)

Each range includes the respondents in the following range or ranges. This was done to match the following possible survey questions from other studies: (a) Do you use this Level of evaluation?, (b) Do you use this Level of evaluation in more than half your programs?, and (c) Do you use this Level of evaluation in most of your programs? Sixty and 80% were used as break points because they matched the breakpoints used in the survey's questions on methods and helps maintain consistency in reporting.

Using 60% and 80% as minimum percentages of programs using a level of evaluation as criteria for concluding that organizations use a level produced numbers that closely match results reported elsewhere (Carnevale & Schulz, 1990; Dixon, 1987; Phillips, 1991; Robinson and Robinson, 1989; Corporate HRD Executive Survey, 1989). Those results that are directly comparable are displayed on Table 2.

**Table 2 - Percentages and Frequencies for Organizations Using Each Level of Evaluation in Some, More than Half, and Most of Their Programs**

| Where reported | Proportion of Programs | Level 1 % | Level 2 % | Level 3 % | Level 4 % |
|---|---|---|---|---|---|
| Current study | Most (>80% of prog.) | 64 | 29 | 15 | 11 |
| | More than 60% | 72 | 42 | 21 | 17 |
| | Some (>0%) | 92 | 84 | 65 | 53 |
| Training Magazine (1996) | any percentage of programs > 0 | 86 | 71 | 65 | 49 |
| Carnevale & Schulz, 1990 | Unknown | 75 to 100 | 100 | 10 | 25 |
| Dixon, 1987 | Unknown | 100 | 30 | 15 | |
| Phillips, 1991 | Unknown | 52 | | 15 & 17 | 8 & 13 |

***Research Question #2 - What impediments exist to the use of evaluation at all levels?*** For each level of evaluation respondents were asked why, on occasions when they did not evaluate at a particular level, did they elect not to evaluate. For each level a series of possible reasons was provided as well as an opportunity to supply a unique reason. Table 3 shows the results, with the top three reasons at each level underlined.

The most commonly reported reason for not evaluating across all levels was that it was not required by the organization (28.83 - 42.34%). The percent citing this reason was highest at levels 3 and 4. A lack of training was cited as one of the top three reasons for levels 2, 3, and 4. Cost was one of the top three reasons for levels 3 and 4.

**Table 3 - Percent Not Evaluating at Each Level**

| Reason | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| Not required | 28.83 | 36.94 | 44.14 | 42.34 |
| Little Value | 18.92 | 19.82 | 13.51 | 15.32 |
| Cost | 10.81 | 18.02 | 36.94 | 36.94 |
| Not legally required | 9.91 | 14.41 | 7.21 | 8.11 |
| Lack of training | 9.01 | 23.42 | 34.23 | 39.64 |
| Time | 8.11 | 21.62 | 3.60 | 5.41 |
| Union | 1.80 | 4.50 | 2.70 | 1.80 |
| Prohibited | .90 | 3.60 | 1.80 | 1.80 |

***Research Question #3 - What methods are used at levels 2, 3, and 4?*** Because space is limited, only summary information can be provided for this question. Use of the various evaluation types or methods was addressed in two different ways. The first measure was the percentage of programs using the method. The second was the amount of use by those organizations that reported using the method. It was therefore possible for a method to be used by a small percentage of the reporting organizations but to have a high percentage of use within those organizations using the method.

The Level 2 evaluation method reported as most used was skill demonstrations. This method was also used in the highest percent of programs by the organizations using the method. The second most common evaluation method at Level 2 was the posttest with no pretest. On-the-job demonstrations, simulations, and work samples were the methods used in the fewest number of programs and organizations. Common use of a method across organizations did not always indicate that a method was used regularly to evaluate programs. Post-tests were reported as used by well over

half the organizations but were reported most commonly used to evaluate only one fifth or less of the programs taught by those organizations.

Observation was reported as the most commonly used Level 3 evaluation method. It was used by slightly over half the reporting organizations. Performance appraisals were the second most common evaluation method and were reported as used by slightly less than half the respondents. However, for the organizations using performance appraisals, they were reported as used in more programs than any other Level 3 method. Other methods commonly reported as used were existing records, records produced for evaluation, action plans, and performance contracts. Surveys and interviews were not commonly used.

Productivity estimates, productivity measures, and regulation compliance were reported as used by one third or more of the organizations for Level 4 evaluation. However, these three methods were only used in a few of the programs taught. ROI measures were used by less than one fifth of the respondents, but those using ROI used it in approximately half the programs taught. All other Level 4 evaluation methods were used in less than one fifth of the programs taught.

***Research Question #4 - Is there a relationship between selected demographics, training practices, and the value placed on training by training managers, and what levels of evaluation are used?*** Exploratory correlations were run between various organizational variables and the four levels of evaluation. Recognizing that with many correlations, relationships may be a function of the number of correlations and not replicable in future studies. Strong correlations would have suggested areas for possible study future study. However, few significant correlations were found and only one exceeded .30 (see Table 4).

**Table 4 - Significant correlations between organizational factors and evaluation practices**

| Organizational factor | Statistically significant correlations greater than .30 |
|---|---|
| business and industry function | — |
| business and industry size | — |
| training manager's perceptions of the importance of a level | (some in the .20 range) |
| training manager's experience | — |
| length of time an organization had been training | — |
| training methods | — |
| training manager's perception of the importance of a level of evaluation in demonstrating value to mgt. | — |
| perceived importance in improving programs | — |
| programs delivered to change performance or org. outcomes | (some in the .20 range) |
| percent of staff involved in evaluation | level 1, $r=.31$, $p<.01$ |
| percent of staff with formal training in evaluation | — |
| percentage of programs dependent on evaluation for funding | — |
| training managers' perceived values for each evaluation level | — |

## Conclusions

This study has confirmed other evaluation practice surveys. Its unique contribution lies in using a rigorous research design, and investigating technical training evaluation which is often thought to be easier to evaluate. By showing that even technical training is not evaluated any more often than general surveys have shown, this study provides even more convincing evidence that evaluation use is a major problem in HRD.

When following the literature back to the development of the four levels, there has been little change in the amount of evaluation within business and industry for at least the last 25 years. For example, Gutek (1988) cited a 1968 study that asked if organizations used each of the four levels. That study reported that 78 % used Level 1, 50 % used Level 2, 54 % used Level 3, and 45 % used Level 4. The usage in the older study is similar to that of this study for Levels 1 and 2 as discussed above but not for Levels 3 and 4. However, the older study stated that the evaluation at Levels 3 and 4 were superficial and subjective. Using a less strict interpretation of use to match the older study ( "used in some of the organization's programs" or "used in more than half of the organization's

programs") this study had a reported use of 21% to 66% for Level 3 and 17% to 55% for Level 4. The range of findings in this study include the values from this 25-year-old study. As early as 1953, Wallace and Twitchell (1953) were discussing the need for and lack of training evaluation. Today's literature contains parallel comments and the lack of training evaluation still exists.

**A New Agenda for Training Evaluation Research?**

It may be time to ask new questions about the use of training evaluation. Television news, news magazines, economic reports, management journals, training literature, politicians, and business leaders all discuss the need to increase the rate of growth in productivity, downsizing, cost-cutting, and ever increasing competitiveness. The training literature presents evaluation as a necessary component in providing training that can help organizations increase productivity, reduce the required number of employees, cut costs, and increase competitiveness. Yet, all the literature on how much evaluation is used by business and industry suggests that less than half of the training programs in place are evaluated for objective outcomes. Additionally, less than one third of training programs are evaluated in any way that measure changes in organizational goals or profitability. It is important to note that these studies all started with a trainer's notion of what constitutes effective evaluation, not a business managers'.

Either business and industry are not concerned with the costs and outcomes of training or they are using some other method of justifying costs and checking outcomes. After at least 40 years of bemoaning the lack of evaluation, supporting the value of evaluation, developing methods of evaluation, and pushing the evaluation cause, there appears to have been little change in the amount, types, or quality of evaluation in business and industry. As in many other endeavors, history keeps repeating itself. Program developers and trainers see the obvious value of evaluation because it would affirm the value of what they do and provide an objective basis for improving both programs and teaching.

However in the face of business and industry's resistance to formal evaluation, those in the training field must ask, is it truly possible that business and industry spend billions of dollars without verifying the value of what they purchase? This is highly unlikely. It is time to ask, "how are business and industry placing value on training?" instead of complaining that they do not evaluate training by the accepted methods reported in the literature. Additionally, it is worth asking why evaluations aren't recognized as a management tool?

Recent research in the use of utility analysis sheds some light on this problem. Utility analysis is a sophisticated technique to determine the dollar value (i.e. utility) of interventions that affect human performance (Cascio, 1991). In two experimental studies, managers were shown data on a new human resource practice (Latham & Wythe, 1994; Wythe & Latham, 1997). Some managers were taught how to interpret the utility analysis, which provided the best cost/benefit data. Control group managers were given other data that was less complete. Based on rational economic theory, it was expected that managers who would seek and use the best cost-benefit data available to make the decision. In both studies, exactly the opposite occurred! After learning about utility analysis, fewer managers preferred to use that data. For some reason, managers moved away from the more sophisticated tools once they learned about them. Statistical techniques such as statistical process control has also struggled to gain widespread utilization, perhaps because they are not strategic in their application (Hoerl, 1995).

This research raises puzzling questions and offers few answers. However, one can not help but wonder if something similar could be happening with training evaluation. Researchers interested in evaluation practices, such as this study, have typically started with some framework or lens that is presumed to be "correct" practice and then investigated the extent to which that framework is being used in practice. In training evaluation, the prescriptive frameworks have generally failed to be confirmed by practice. The usual interpretation has been that practitioners aren't doing what they should. An alternate interpretation may be that the models are not correct.

New research is needed that takes a more objective look at the decision and influence

processes that managers use to make decisions about training. Such research should not start with a prescriptive framework, but rather study actual decision processes in action. Likely research questions would include:

1. What criteria are used to determine:
   - whether training should be initiated?
   -- whether training was effective?
   -- whether training changed performance?
   -- whether to continue using a form of training?
   – whether to adopt a different method?
   – whether to "buy" training services from a "seller"
2. How are best practices determined?
3. Who are the opinion leaders that most influence training decisions?
4. What are the actual decision steps used?
5. What factors or events lead to changes in training practices?

**References**


Brinkerhoff, R. O. (1991). Achieving results from training. San Francisco: Jossey-Bass.

Broad, M. L. & Newstrom, J. W. (1992). Transfer of training; action-packed strategies to ensure high payoff from training investments. Reading, MA.: Addison-Wesly.

Carnevale, A. P., & Schulz E. (1990). Economic accountability for training: Demands and responses. Training and Development Journal, 44(7), s-1 - s-32.

Cascio, W. F. (1991). Costing human resources: The financial impact of behavior in organizations. Boston: PWS-Kent Publishing Co.

Corporate HRD Executive Survey. (1989). Survey #11 Report. Alexandria, VA: American Society for Training and Development.

Dillman, D. D. (1978). Mail and telephone surveys: The total design method. New York: John Wiley & Sons. Inc.

Dixon, N. M. (1990). Evaluation: a tool for improving quality. San Diego: University Associates, Inc.

Dixon, N. M. (1987). Meeting training's goals without reaction forms. Personnel Journal, 66(8) 108-115.

Gordon, J. (1991, August). Measuring the goodness of training. Training, 30(8), 19-25.

Gutek, S. P. (1988). Training program evaluation: An investigation of perceptions and practices in non-manufacturing business organizations. Dissertation Abstracts International, 49/05A, 1040.

Hoerl, R. W. (1995) Enhancing the bottom-line impact of statistical methods. Quality Management Journal. Summer, 58-69.

Holton, E. F. (1996). The flawed four level evaluation model. Human Resource Development Quarterly, 7,

Industry report 1996. (1996, October). Training, 33(10), 36-49.

Kirkpatrick, D. L. (1994). Evaluating training programs: the four levels. San Francisco: Berrett-Koehler.

Latham, G. & Wythe, G. (1994). The futility of utility analysis , Personnel Psychology, 47, 31-46.

Phillips, J. J. (1991). Handbook of training evaluation and measurement methods. Houston: Gulf.

Robinson, D. G.& Robinson, J. C. (1989) Training for impact: How to link training to business needs and measure the results. San Francisco: Jossey-Bass.

Wallace, S. R. Jr.& Twitchell, C. M. (1953). Evaluation of a Training Course for Life Insurance Agents. Personnel Psychology, 6(1) 25-43.

Wythe, G. & Latham, G. (1997). The futility of utility analysis revisited: when even an expert fails, Personnel Psychology, 50, 601-610.